2014 DLF Forum

An All Purpose Archives Viewer: Displaying Large Scale Archival Collections in Digital Libraries

Monday, October 27, 10:45am-12:00pm
Salons 4,5,6, Georgia Tech Hotel and Conference Center

Colocated with: **Spotlight: A Self-Service Tool for Showcasing Digital Collections** (Group notes: https://docs.google.com/document/d/1ZjUskMQ305v2mHL9jx-_PfrD68W2BRiTFy6Zvr_24mY/edit#)

Presenters
Elizabeth McAulay, University of California, Los Angeles
Kristian Allen, University of California, Los Angeles

Notes

Looking at large scale digitization of archives. “Scanning everything” fits archival practice. The whole rather than the constituent parts is key. Respecting the entirety of the archive. How to embrace less process more product.
Using finding aid to provide metadata; experimented with mapping to MODS - but not satisified Didn’t need to crosswalk, had everything needed with EAD (sufficient metadata, structure). Adding item level links to EAD, maintaining URLs; created disjointed, item-level approach/experience.End result was Online Archive of California. Trying to make system more integrated. Looked at Archives of American Art as exemplar. AAA has a custom built solution; UCLA looking at how to do it for themselves; UNC another exemplar (Southern Historical Collection). Use a file naming scheme that captures container information - connect to file with little additional work. Getting the components to talk to each other. UCLA had a lot of in-house Drupal expertise and local SOLR and Fedora knowledge. Also implementing Islandora as backend to run projects; expose front-end via Fedora. Discovery Garden used for their expertise for customizing Islandora/Fedora pieces. Three main components: EAD processing (platform agnostic, XSLT, Javascript-based); Browse and Discovery (apache solr, drupal, fedora); Archival Viewer (tei, seadragon). “agnostish” XSLT can be used independently of Islandora and could be customized.
EAD Tree Map: apply XSLT to EAD to generate static HTML map (and feed into Drupal) or CSS or other means of styling… Using fieldset tag. Links to Solr queries to get contents of box or folder in fedora. Solr provides the capability to link EAD to content; associate a node with box or folder that you want results from. Number one - get huge amounts of content in, and then double-back and flesh out the metadata.
Archival Viewer Component: TEI and image side by side; example: mss image and text

loaded a stock Islandora viewer for demo. 1) load EAD and see tree map that is generated.; need to associate uploaded assets with a Solr query...uses an interface to associate items with EAD, uses a dynamic Javascript tree that you use to locate your folder/content to associate/attach. This allows the finding aid to locate results tied to the item.
This is first draft; next is lots of automation - linking to solr queries is still manual, want to write script to generate link based on file name. Isadora Duncan collection releasing this fall; initial example. About 8000 objects.

Questions: Has this process changed how the EAD finding aids are written? (Not yet, speakers hopes it doesn't…)
relationship with OAC - continue to be repository for finding aids; integrating with digital asset development. Ability to add more metadata at item-level, each component is individually managed.

Within the archives world, it is entirety of the collection not single items.
How do we make collections available, MPLP as our dominant mantra
UCLA wanted to embrace this ethos; use the finding aid to provide the metadata. The EAD allowed the data to be cross-walked.
Archivists considered use and access from the patron's perspective. OAC links from locally produced EAD to full-text content or images. Still not perfect. Exemplars include UNC, Chapel Hill, Smithsonian Archives of American Art.
EAD is a great digital tool, but it is necessary that it also connects to the physical item (box numbers, etc)

Kristian: UCLA uses Drupal in-house expertise, Local Solr and Fedora knowledge, implements Islandora as backend, Discovery Garden as tech solution

3 main components today are platform agnostic
1) EAD processing EAD=XSLT create HTML "tree map"
2) Browse and Discovery Solr query used to find related assets
3) Archival Viewer Javascript used

EAD tree map:
XSLT applied to any valid EAD
Navigable HTML using fieldset tag
Links to leaf items are Solr queries

First step is to get huge amounts of data, then return and flush out the details

Archival Viewer Component
Based on OpenSeadragon
View TEI and image side-by-side

Key Features; TEI and image synchronized updates, Crop, save, and vew

Here is where the team does a visual archive viewer demo.

Final thoughts and next steps
First cut, room for improvement
automation is the next and major step
Isadora Duncan launch this fall. (entire collection is digitized)

Link to spotlight notes:
https://docs.google.com/document/d/1ZjUskMQ305v2mHL9jx-_PfrD68W2BRiTFy6Zvr_24mY/edit?pli=1